

Pozytywne prymowanie semantyczne jako
narzędzie optymalizacji automatycznych
systemów autoryzacji użytkownika

RAPORT Z BADAŃ

Dr Mariusz Urbański
Zakład Logiki i Kognitywistyki
Instytut Psychologii
Uniwersytet im. A. Mickiewicza

Mgr Paweł Łupkowski
Zakład Logiki i Kognitywistyki
Instytut Psychologii
Uniwersytet im. A. Mickiewicza

STRESZCZENIE

Naszym głównym celem było sprawdzenie hipotezy głoszącej, że pozytywne prymowanie semantyczne wystąpi przy pewnego typu zadaniach leksykalnych i w znaczny sposób poprawi czas ich rozwiązania. Potwierdzenie tej hipotezy ma bardzo duże znaczenie dla projektowanego przez nas automatycznego systemu autoryzacji użytkownika w środowisku WWW.

Spis treści

1	Wprowadzenie	4
1.1	Problem	4
1.2	CAPTCHA	5
1.3	Projekt badawczy	6
2	Metoda	8
2.1	Konstrukcja zadań testowych	8
2.2	Osoby uczestniczące w badaniu	10
2.3	Plan i przebieg badania	13
2.4	Sposób analizowania wyników	14
3	Rezultaty	15
3.1	Czas rozwiązania zadań	15
3.2	Poprawność rozwiązania zadań	18
3.3	Trudność zadań	22
3.4	Użyteczność	23
4	Dyskusja	25
5.	Literatura	26
6.	Załączniki	29

1 Wprowadzenie

1.1 Problem

Powszechny dostęp do globalnej sieci Internet stanowi niewątpliwie ogromny krok naprzód w dziedzinie komunikacji, jednocześnie jednak stwarza równie ogromny problem zapewnienia bezpieczeństwa owej komunikacji. Jedną z głównych zalet Internetu — szeroko zakrojona automatyzacja komunikacji — jest również jednym z najbardziej oczywistych i najgoręcej dyskutowanych zagrożeń: korzystając z Internetu chcielibyśmy mieć pewność, że dane, które przesyłamy lub udostępniamy w globalnej sieci nie zostaną wykorzystane przez osoby do tego niepowołane. Dotyczy to przede wszystkim następujących dziedzin wymiany i udostępniania danych w sieci WWW (por. [5], [8], [12], [14]).

- *Darmowe konta e-mailowe.* Specjalnie skonstruowane programy (*boty*) rejestrują tysiące darmowych kont, aby później rozsyłać z nich spam lub wirusy komputerowe.
- *Serwisy udostępniające darmowe usługi.* Część firm oferuje nieodpłatne korzystanie z pewnych usług, np. wyszukiwarek, ofert, katalogów a także z meta-usług, takich jak programy umożliwiające porównywanie cen u różnych producentów. Dostarczenie takich usług jest dla firmy kosztowne, ale spełnia swój konkretny marketingowy cel: przyciąga i wiąże klienta z daną firmą. Sytuacja ta zmienia się, gdy ktoś wykorzystuje — dostarczane nieodpłatnie — dane i usługi dla własnych korzyści (głównie do tzw. *comparative shopping*).
- *Serwisy, w których liczy się liczba przeprowadzanych przez użytkownika transakcji, wizyt itp.* Jeżeli pozycja i uprawnienia użytkownika zależą od pewnego wskaźnika, którym jest, powiedzmy, liczba wizyt na pewnej stronie WWW, to chcielibyśmy mieć pewność, że użytkownik ten jest konkretną osobą, nie zaś *botem*.
- *Prywatność i ograniczony dostęp do danych.* Serwisy oferujące możliwość zabezpieczenia danych *loginem* i *hasłem* powinny mieć możliwość rozpoznania prób skorzystania z takich danych przez niepowołanego użytkownika (np. poprzez atak typu *Denial-Of-Service* lub *Dictionary Attacks*).

Jak łatwo zauważyć, wymienione zagadnienia wiążą się z problemem automatycznego rozpoznania, czy system ma do czynienia z człowiekiem, czy ze specjalnie skonstruowanym programem komputerowym. Taki właśnie cel stawia się przed systemami CAPTCHA.

1.2 CAPTCHA

CAPTCHA (*Completely Automated Public Turing Test to Tell Computers and Humans Apart*) jest protokołem kryptograficznym, którego konstrukcja oparta jest na pewnym trudnym problemie z dziedziny sztucznej inteligencji (SI).

Najistotniejsze cechy CAPTCHA (por. [3], [2], [14]):

1. Kolejne instancje testu są generowane automatycznie.
2. Test może zostać szybko rozwiązany przez użytkownika, który jest człowiekiem.
3. Test przechodzą jedynie ludzie (bez konieczności odwoływania się do specjalistycznej wiedzy).
4. Testu nie są w stanie zdać automaty (*boty*).
5. Test pozostaje aktualny, pomimo zmian w technologii i metodach programistycznych (pomimo publicznego dostępu do algorytmów owego testu).

Zalety CAPTCHA:

- CAPTCHA nie bazują na trudnych obliczeniowo problemach z teorii liczb ale na otwartych i trudnych problemach z dziedziny SI, dzięki czemu:
 - są bardziej intuicyjne dla użytkownika,
 - są wygodniejsze w stosowaniu (autoryzacja użytkownika odbywa się bez konieczności zapamiętywania haseł itp.),
 - stanowią motywację dla badaczy z dziedziny SI (por. [13]).
- Wdrożenie takich systemów jest stosunkowo łatwe i tanie.

Warto zwrócić uwagę, że cechy jakimi charakteryzuje się systemy CAPTCHA pozwalają na skonstruowanie takiego systemu w oparciu o wyniki badań nie tylko z dziedziny informatyki ale również psychologii bądź kognitywistyki etc.

Źródłem inspiracji dla takiego systemu może być — naszym zdaniem — technika eksperymentalna rozwinięta na gruncie psychologii poznawczej, a mianowicie *prymowanie*.

Najogólniej mówiąc technika prymowania (poprzedzania) opiera się na eksponowaniu bodźców poprzedzających przed eksponowaniem bodźca właściwego (por. [21, s. 795]).

W kontekście naszej propozycji szczególnie interesujące jest zjawisko prymowania semantycznego. Pryma nie jest tu identyczna z bodźcem, ale jest z nim *znaczeniowo* powiązana (np. biblioteka → księgozbiór), lub należy do tej samej *kategorii* (np. długopis → ołówek).

W badaniach tego typu stwierdza się zazwyczaj, że bodziec prymujący ułatwia przetwarzanie bodźca docelowego, np. skraca czas potrzebny na jego rozpoznanie, przeczytanie, nazwanie itp. Mamy wtedy do czynienia z **poprzedzaniem pozytywnym**, zwanym też niekiedy torowaniem, ponieważ pryma jak gdyby toruje drogę bodźcowi docelowemu. [15, s. 225]

1.3 Projekt badawczy

Nasz projekt podejmował zatem problematykę pozytywnego prymowania semantycznego. Celem naszym było zbadanie zależności czasu rozwiązania zadania testowego w zależności od występowania prymy lub jej braku.

Inspirację dla podjęcia projektu stanowiła analiza istniejących systemów, służących automatycznemu rozróżnianiu pomiędzy użytkownikiem, który jest człowiekiem, a *botem* oraz analiza zjawiska pozytywnego prymowania semantycznego.

Do podjęcia proponowanego tematu skłoniły nas przesłanki *poznawcze* (zagadnienie możliwości występowania zjawiska prymowania w przypadku zadania polegającego na wyszukiwaniu różnic) oraz przesłanki *praktyczne* (zaprojektowanie i zaimplementowanie systemu informatycznego opartego na wynikach badań dotyczących prymowania).

Postawione zostały następujące pytania badawcze:

1. Czy w przypadku określonego typu zadań możemy mówić o pozytywnym prymowaniu semantycznym?
2. Czy istnieje średni czas rozwiązania takiego typu zadania dla interesującej nas populacji?
3. Jakie zalety ma proponowany system w porównaniu z istniejącymi rozwiązaniami (pod względem: użyteczności, bezpieczeństwa)?

W związku z powyższymi pytaniami sformułowane zostały następujące hipotezy badawcze:

- H1 W przypadku określonego typu zadań testowych wystąpi efekt pozytywnego prymowania semantycznego.

H2 Średni czas rozwiązania określonego typu zadań testowych pozwoli różnicować grupę eksperymentalną od kontrolnej (grupa eksperymentalna będzie rozwiązywała zadania szybciej).

H3 W przypadku grupy eksperymentalnej deklarowana subiektywna trudność zadań będzie niższa od deklarowanej przez grupę kontrolną.

Aby zweryfikować sformułowane hipotezy badawcze wyróżniono następujące zmienne do zbadania w eksperymencie:

- *występowanie / brak prymowania w przypadku proponowanego zadania* (zmienna niezależna),
- *średni czas rozwiązywania zadania testowego* (zmienna zależna),
- *subiektywna miara trudności zadania testowego* (zmienna zależna).

2 Metoda

2.1 Konstrukcja zadań testowych

Dla potrzeb eksperymentu zaprojektowano dziesięć zadań. Każde z nich prezentowane było w postaci kolorowego (w trybie RGB) obrazka o *wymiarach 227 / 142 px* i *rozdzielczości 72 dpi*. Na pojedyncze zadanie składały się trzy słowa (będące nazwami zwierząt). Użyto tutaj *czcionek: Bitstream Charter Italic* oraz *Century Schoolbook L, Italic*. Jedno ze słów różniło się od pozostałych (będąc nazwą zwierzęcia innego *rzędu* niż pozostałe dwa).

Przygotowano dwa zestawy zadań różniące się tylko tym, że w jednym z nich każde zadanie poprzedzone było prymą znaczeniowo powiązaną ze słowem, które stanowiło rozwiązanie owego zadania. W przypadku prymy wykorzystano *czcionkę: Bitstream Sans Sherif*.

Poniżej prezentujemy zestawy słów wykorzystanych w poszczególnych zadaniach (słowa w nawiasach to prymy):

1. (*jaszczurka*) baran—tarpan—waran;
2. (*ryba*) koń—słoń—okoń;
3. (*dziób*) żbik—dzik—kruk;
4. (*małpa*) murena—makrela—makak;
5. (*pióro*) krowa—lama—wrona;
6. (*dziób*) pies—kot—kos;
7. (*kopiec*) kura—kruk—kret;
8. (*mleko*) kaczka—kukułka—krowa;
9. (*źródło*) koza—owca—osa;
10. (*mysz*) kruk—kos—kot.

Dodatkowo każde z zadań (pomijając prymy) zostało poddane zniekształceniom (tak aby utrudnić rozpoznanie słów algorytmom OCR — *Optical Character Recognition*). Wykorzystano tutaj standardowy zestaw zniekształceń stosowanych dla CAPTCHA: rozmycie Gaussa, rozmycie HSV, dodanie mgły, rozrzucenie tekstu, W&P, wydruk gazetowy. Poniższe zestawienie zawiera poszczególne zniekształcenia wraz z ich parametrami (zniekształcenia wykonano przy użyciu programu GIMP — v.2.2.10):

- **Rozmycie Gaussa (G-blur)**
IIR; promień: 5,0 / 5,0 px
- **Rozmycie HSV (HSV)**
wstrzymanie: 2
odcień: 3
nasycenie: 22
wartość: 22
- **Dodanie mgły (fog)**
(Python-Fu/effects/add Fog)
typ: Clouds
turbulence: 1,0
opacity: 100,0
- **Rozrzucenie tekstu (rozzr.)**
Rozmiar rozrzutu: 5 / 5 px
- **Whirl and Pinch (W& P)**
(Python-Fu/disorts/Whirl and Pinch)
Whirl angle: 90
Radius: 1
- **Wydruk gazetowy (wydr. gaz.)**
SPI: 72
CPI: 7,2
Rozmiar oczka: 10
Rozdzielane kanały: RGB

Zestaw zadań testowych poddanych zniekształceniom znajduje się w **Załączniku A**. Efektywność wprowadzonych zniekształceń została wypróbowana przy użyciu jednego z lepszych programów do rozpoznawania pisma: *AABYY Fine Reader PE 8.0*. Efekty tej analizy przedstawione są w **Załączniku B**.

Każde z zadań testowych posiadało indywidualny zestaw charakterystyk, obejmujących:

1. czas wyświetlania prymy,
2. czas wyświetlania maski,
3. metodę zniekształcenia tekstu,
4. metodę zniekształcenia tła.

Poniższa tabela prezentuje wspomniane charakterystyki zadań testowych.

Tab.1. Charakterystyki zadań testowych

nr	pryma (ms)	maska (ms)	znieksz. tekstu	znieksz. tła
T1	70	50	G-blur	HSV
T2	60	50	G-blur	RGB
T3	80	50	G-blur	fog
T4	90	50	rozrz.	HSV
T5	100	60	rozrz.	RGB
T6	60	30	rozrz.	fog
T7	70	50	W& P	HSV
T8	70	50	W& P	fog
T9	70	50	W& P	RGB
T10	70	50	wydr. gaz.	HSV

Zadania testowe przygotowane zostały jako obrazki w formacie *gif*, co miało swoje znaczenie dla zadań prymowanych (gwarantowało odpowiednie czasy wyświetlania prymy i maski). Na potrzeby badania skonstruowano specjalną stronę internetową, dzięki czemu uzyskano dużą automatyzację przebiegu badania począwszy od mierzenia czasu rozwiązania zadania, na wskazywaniu poprawności rozwiązania skończywszy.

Dodatkowo — w celu wstępnego zbadania aspektów użytecznościowych systemu, oraz uzyskania podstawowych informacji o badanych — skonstruowano krótką ankietę, której wzór znajduje się w **Załączniku C**.

2.2 Osoby uczestniczące w badaniu

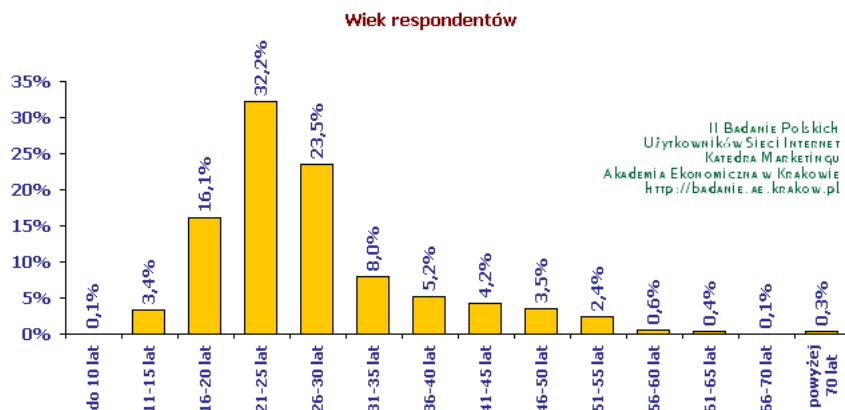
Ogółem przebadano **64** osoby. W sposób losowy rozdzielono badanych między dwie grupy: *A* — grupę z zadaniami prymowanymi; *B* — grupę z zadaniami bez prymowania. W grupie *A* znalazło się 31 osób, zaś w grupie *B* 33 osoby.

Tab.2. Podział osób badanych na grupy

	Częstość	Procent
<i>A</i>	31	48,4
<i>B</i>	33	51,6
Ogółem	64	100,0

Średnia wieku w grupie *A* to 21,52 zaś w grupie *B* 21,03. Taki dobór wieku osób badanych motywowany jest wynikami badania polskich użytkowników Internetu. Badania przeprowadzone przez Akademię Ekonomiczną w

Krakowie (II Badanie Polskich Użytkowników Sieci Internet – 2000) wskazują, że 32 % badanych użytkowników Internetu to osoby w wieku 21-25 lat¹. Dane te potwierdzają badania firmy I-Metria z roku 2004².



Rys.1. Użytkownicy Internetu w Polsce ze względu na wiek

Grupa *A* liczyła 11 mężczyzn i 18 kobiet (2 osoby – brak danych), zaś grupa *B* liczyła 8 mężczyzn i 25 kobiet.

Jeżeli chodzi o wykształcenie, to w grupie *A* 83,9 % stanowiły osoby z wykształceniem średnim, 16,1 % osoby z wykształceniem wyższym. W grupie *B* osoby z wykształceniem średnim stanowiły 87,9 %, zaś osoby z wykształceniem wyższym 12,1 %.

Tab.3. Grupa badana – wykształcenie

grupa	wykszałcenie	częstość	procent
<i>A</i>	średnie	26	83,9
	wyższe	5	16,1
	ogółem	31	100,0
<i>B</i>	średnie	29	87,9
	wyższe	4	12,1
	ogółem	33	100,0

Zgromadzono również dwie dodatkowe charakterystyki dotyczące grupy badanej, które są szczególnie istotne ze względu na cel badań:

- częstotliwość korzystania z Internetu,

¹<http://www.badanie.ae.krakow.pl>

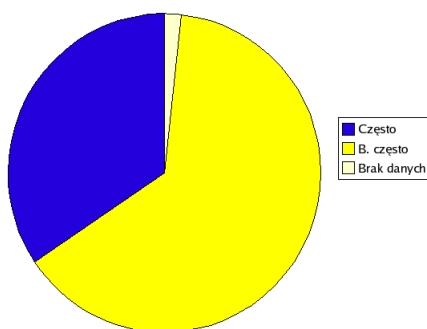
²<http://www.cxo.pl/news/44192.html>

- korzystanie z darmowych kont e-mailowych.

W przypadku częstotliwości korzystania z Internetu badani deklaruwali jedną z trzech odpowiedzi:

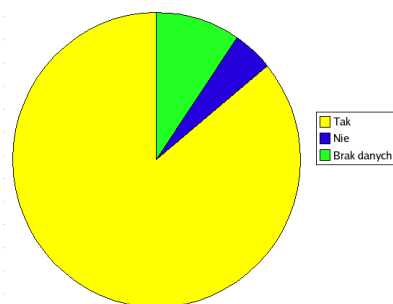
- okazjonalnie,
- często (*przynajmniej raz w tygodniu*),
- bardzo często (*codziennie*).

Charakterystyka grupy badanej ze względu na wymienioną zmienną przedstawia się następująco: 3 na 64 osoby zadeklarowały jedynie *okazjonalne* korzystanie z Internetu; 21 na 64 *często*; zaś 39 *bardzo często* (w przypadku 1 osoby badanej nie uzyskano danych).



Rys.2. Jak często korzystasz z Internetu?

Jeśli chodzi o korzystanie z darmowych skrzynek e-mailowych, to aż 55 badanych potwierdziło ich używanie (3 osoby nie korzystają z darmowych skrzynek e-mail, w przypadku 6 osób nie udało się zebrać danych).

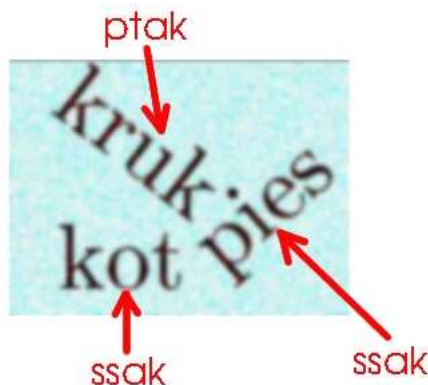


Rys.3. Czy korzystasz z darmowych skrzynek e-mail?

2.3 Plan i przebieg badania

Główna część badania poświęcona była rozwiązywaniu zadań testowych. Badani przy wchodzeniu do sali, w której odbywało się badanie byli losowo rozdzielani do jednego z 8 stanowisk komputerowych (a tym samym przydzielani losowo do jednej z grup: *A* (grupa eksperymentalna z zadaniami prymowanymi) lub *B* (grupa kontrolna). Wszystkie stanowiska komputerowe posiadały takie same monitory z ustawioną identyczną rozdzielczością ekranu. Zadania testowe umieszczone były na specjalnie przygotowanej stronie WWW, co zapewniało ich identyczność na każdym ze stanowisk.

Osoba badana najpierw wypełniała krótki formularz dotyczący jego danych, po czym podawane zostały instrukcje dotyczące rozwiązywania zadań i badanym prezentowane było przykładowe zadanie wraz z rozwiązaniem (zawsze przez jedną osobę spośród prowadzących eksperyment— tak aby zapewnić spójność udzielanych instrukcji).



Rys.4. Przykładowe zadanie testowe

Następnie badany rozwiązywał kolejne 10 zadań (czas mierzony był automatycznie przez serwer). Podanie odpowiedzi przez badanego odbywało się poprzez wskazanie kursorem i kliknięcie myszką słowa, które badany uważał za rozwiązanie. Powodowało to zatrzymanie wyświetlania zadania testowego, na ekranie zaś pojawiał się czas rozwiązania oraz zakodowana informacja o poprawności rozwiązania.

Po rozwiązaniu ostatniego zadania badany wypełniał krótką ankietę, w której określał subiektywny poziom trudności poszczególnych zadań (po zakończeniu rozwiązywania zadań na ekranie pojawiała się ich zestawienie), określając ją na skali od 1 (najłatwiejsze) do 10 (najtrudniejsze). Badany odpowiadał również na pytania o ewentualną możliwość rozwiązywania tego

typu zadań przy pewnych procedurach związanych z użytkowaniem Internetu.

Całość przebiegu badania nadzorowana była przez badacza asystującego osobie badanej. Do zadań badacza należało kontrolowanie przebiegu badania, spisywanie czasów rozwiązania zadania podawanych przez serwer, zapisywanie, czy rozwiązanie było poprawne, czy błędne oraz objaśnienie zasad wypełniania krótkiej ankiety.

2.4 Sposób analizowania wyników

Analizę wyników przeprowadzono przy użyciu programu SPSS (12.PL).

Wyniki dotyczące różnicy w średnich czasach rozwiązania zadania przez grupę eksperymentalną i grupę kontrolną zanalizowane zostały z wykorzystaniem testu ANOVA.

Do analizy wyników dotyczących poprawności rozwiązań w obu grupach wykorzystano test χ^2 .

Analiza wyników mówiących o deklarowanym subiektywnym poczuciu trudności zadania przeprowadzona została z użyciem testu t-Studenta.

3 Rezultaty

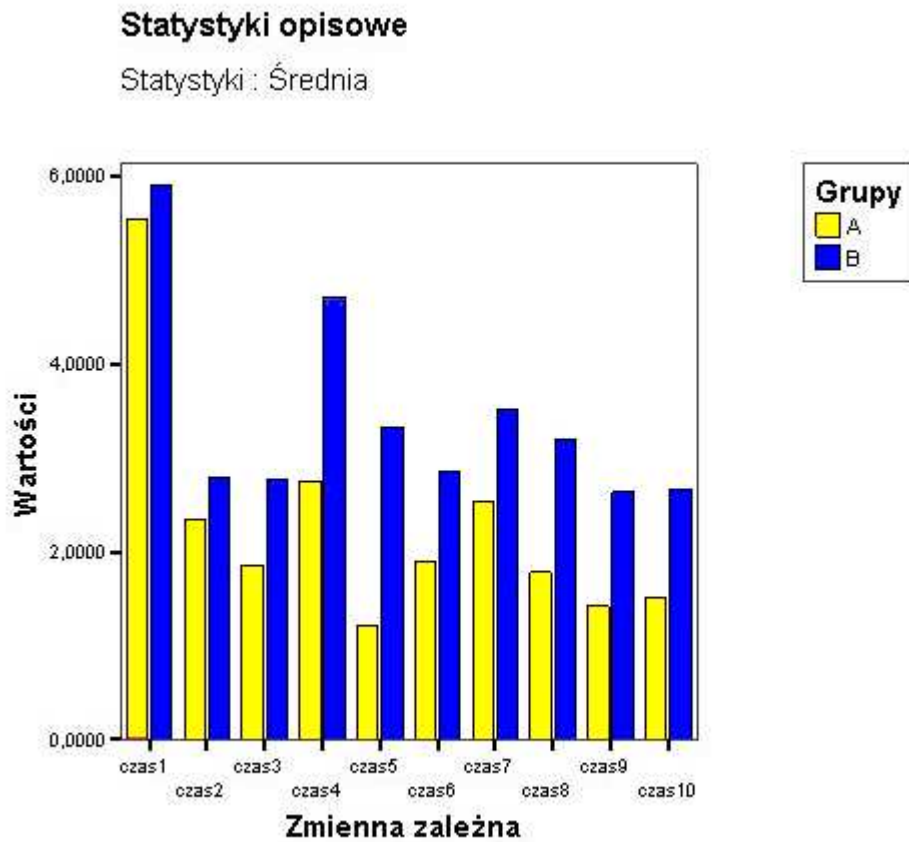
3.1 Czas rozwiązania zadań

Poniższa tabela (Tab.4.) prezentuje średni czas rozwiązania dla poszczególnych zadań z uwzględnieniem podziału na grupy (gdzie *A* — grupa eksperymentalna, *B* — grupa kontrolna).

Tab.4. Średni czas rozwiązania dla poszczególnych zadań

Zadanie	Grupa	N	Średni czas rozwiązania (sek.)	Odchylenie standardowe	Minimum	Maksimum
1.	<i>A</i>	31	5,5408	4,58350	,80	23,77
	<i>B</i>	33	5,9048	3,47367	1,49	17,21
2.	<i>A</i>	31	2,3467	2,22950	,63	9,88
	<i>B</i>	33	2,7859	1,44047	1,36	8,30
3.	<i>A</i>	31	1,8594	1,42309	,42	7,87
	<i>B</i>	33	2,7749	1,54021	1,22	10,19
4.	<i>A</i>	31	2,7456	,92263	,33	4,20
	<i>B</i>	33	4,7085	1,68363	1,69	9,33
5.	<i>A</i>	31	1,2047	,57816	,48	2,73
	<i>B</i>	33	3,3308	2,88513	1,37	18,52
6.	<i>A</i>	31	1,8863	,57865	,79	3,01
	<i>B</i>	33	2,8534	1,01848	1,34	5,80
7.	<i>A</i>	31	2,5314	1,78107	,32	9,63
	<i>B</i>	33	3,5239	1,77903	1,25	10,77
8.	<i>A</i>	31	1,7810	,84231	,81	4,12
	<i>B</i>	33	3,2051	,93643	1,87	5,39
9.	<i>A</i>	31	1,4193	,51846	,56	2,27
	<i>B</i>	33	2,6340	,99129	1,11	6,40
10.	<i>A</i>	31	1,5180	,72901	,07	3,61
	<i>B</i>	33	2,6648	,96555	1,26	5,66

Zależności pomiędzy średnimi czasami rozwiązania zadania dla grupy prymowanej i nieprymowanej przedstawia Rys.5. Żółtym kolorem oznaczono tutaj wyniki grupy eksperymentalnej, niebieskim zaś wyniki grupy kontrolnej.



Rys.5. Średni czas rozwiązania dla poszczególnych zadań

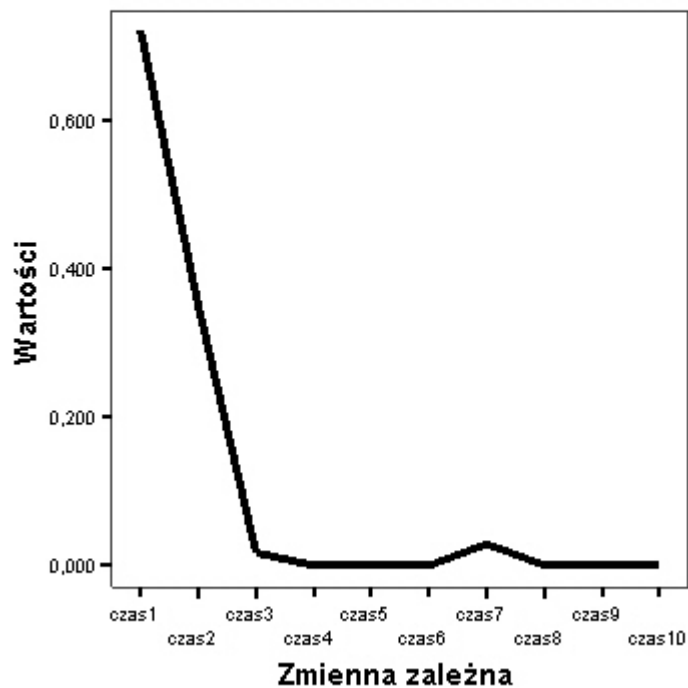
Wyniki jednoczynnikowego testu ANOVA (między grupami) przedstawione są w Tab.5. Czcionką pogrubioną zaznaczono wyniki istotne. Rysunek 6 obrazuje istotność wyników z tabeli 5.

Tab.5. ANOVA dla średniego czasu rozwiązania poszczególnych zadań

Zadanie	Suma kwadratów	df	Średni kwadrat	F	Istotność
zadanie 1	2,118	1	2,118	,129	,721
zadanie 2	3,083	1	3,083	,887	,350
zadanie 3	13,198	1	13,198	5,995	,017
zadanie 4	61,586	1	61,586	32,847	,000
zadanie 5	72,251	1	72,251	16,207	,000
zadanie 6	14,948	1	14,948	21,434	,000
zadanie 7	15,746	1	15,746	4,970	,029
zadanie 8	32,415	1	32,415	40,729	,000
zadanie 9	23,585	1	23,585	37,011	,000
zadanie 10	21,024	1	21,024	28,475	,000

Jednoczynnikowa ANOVA

Źródło zmienności : Między grupami
Statystyki : Istotność



Rys.6. Istotność wyników (na podst. Tab.5.)

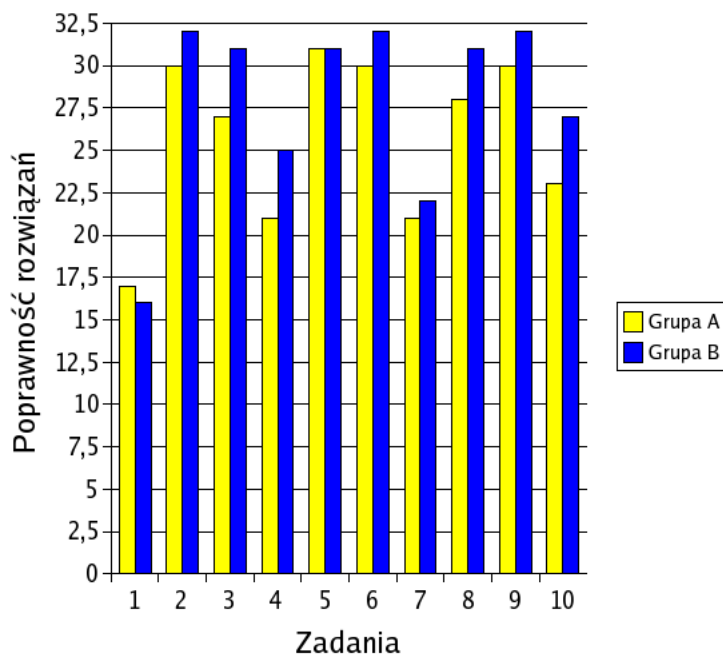
3.2 Poprawność rozwiązania zadań

Sprawdzono również poprawność rozwiązania zadań w zależności od działania eksperymentalnego. Tabela 6 stanowi zestawienie sumy rozwiązań poprawnych dla grupy eksperymentalnej i dla grupy kontrolnej.

Tab.6. Poprawność rozwiązań — porównanie grup

Zad.	Grupa	Poprawne
		Σ
1.	A	17
	B	16
2.	A	30
	B	32
3.	A	27
	B	31
4.	A	21
	B	25
5.	A	31
	B	31
6.	A	30
	B	32
7.	A	21
	B	22
8.	A	28
	B	31
9.	A	30
	B	32
10.	A	23
	B	27

Rysunek 7 przedstawia porównanie sum poprawnych rozwiązań dla poszczególnych zadań uzyskanych dla grupy eksperymentalnej (grupa *A*) i grupy kontrolnej (grupa *B*).



Rys.7. Poprawność rozwiązań — porównanie grup

Poniższe tabele (Tab. 7-16) przedstawiają zestawienie odpowiedzi poprawnych i niepoprawnych dla obu grup (z podziałem na zadania) oraz wyniki testu χ^2 . Czcionką pogrubioną wyróżniono wyniki istotne.

Tab.7. Poprawność rozwiązań — zadanie 1

		A	B
Niepoprawnie	Liczebność	14	16
	% z grupy	45,2 %	50,0 %
Poprawnie	Liczebność	17	16
	% z grupy	54,8 %	50,0 %

$\chi^2=0,15$, $df=1$, $p=0,701$

Tab.8. Poprawność rozwiązań — zadanie 2

		A	B
Niepoprawnie	Liczebność	1	0
	% z grupy	3,2 %	0%
Poprawnie	Liczebność	30	32
	% z grupy	96,6 %	100,0 %

$\chi^2=1,05$, $df=1$, $p=0,306$

Tab.9. Poprawność rozwiązań — zadanie 3

		A	B
Niepoprawnie	Liczebność	4	0
	% z grupy	12,9 %	0 %
Poprawnie	Liczebność	27	31
	% z grupy	87,1 %	100,0 %

$\text{Chi}^2=4,28$ $\text{df}=1$, **$p=0,039$**

Tab.10. Poprawność rozwiązań — zadanie 4

		A	B
Niepoprawnie	Liczebność	9	6
	% z grupy	30,0 %	18,8 %
Poprawnie	Liczebność	21	25
	% z grupy	70,0 %	78,1 %

$\text{Chi}^2=1,89$ $\text{df}=1$, $p=0,390$

Tab.11. Poprawność rozwiązań — zadanie 5

		A	B
Niepoprawnie	Liczebność	0	1
	% z grupy	0 %	3,1 %
Poprawnie	Liczebność	31	31
	% z grupy	100,0 %	96,9 %

$\text{Chi}^2=0,98$ $\text{df}=1$, $p=0,321$

Tab.12. Poprawność rozwiązań — zadanie 6

		A	B
Niepoprawnie	Liczebność	30	32
	% z grupy	100,0 %	100,0 %

Tab.13. Poprawność rozwiązań — zadanie 7

		A	B
Niepoprawnie	Liczebność	10	10
	% z grupy	32,3 %	31,3 %
Poprawnie	Liczebność	21	22
	% z grupy	67,7 %	68,8 %

$\text{Chi}^2=0,01$ $\text{df}=1$, $p=0,932$

Tab.14. Poprawność rozwiązań — zadanie 8

		A	B
Niepoprawnie	Liczebność	3	1
	% z grupy	9,7 %	3,1 %
Poprawnie	Liczebność	28	31
	% z grupy	90,3 %	96,9 %

$\text{Chi}^2=1,14$ $\text{df}=1$, $p=0,286$

Tab.15. Poprawność rozwiązań — zadanie 9

		A	B
Niepoprawnie	Liczebność	30	32
	% z grupy	100,0 %	100,0 %

Tab.16. Poprawność rozwiązań — zadanie 10

		A	B
Niepoprawnie	Liczebność	6	4
	% z grupy	20,7 %	12,9 %
Poprawnie	Liczebność	23	27
	% z grupy	79,3 %	87,1 %

$\text{Chi}^2=0,65$ $\text{df}=1$, $p=0,419$

3.3 Trudność zadań

Zbadano również deklarowaną subiektywną trudność rozwiązywanego zadania w zależności od działania czynnika eksperymentalnego.

Tabela 17 przedstawia porównanie średniej deklarowanej trudności zadania dla grupy kontrolnej i grupy eksperymentalnej z podziałem na poszczególne zadania.

Tab.17. Średnia trudność dla poszczególnych zadań - porównanie grup

zadanie	grupa	N	Średnia	Odchylenie standardowe	Błąd standardowy średniej
zad.1	A	31	6,35	2,973	,534
	B	31	6,55	2,850	,512
zad.2	A	31	2,61	2,290	,411
	B	32	3,56	2,242	,396
zad.3	A	31	3,26	2,250	,404
	B	31	3,48	2,096	,376
zad.4	A	31	5,61	2,765	,497
	B	32	6,50	2,328	,412
zad.5	A	30	3,00	2,017	,368
	B	32	3,63	2,121	,375
zad.6	A	30	3,03	2,157	,394
	B	32	3,47	1,900	,336
zad.7	A	30	4,50	2,570	,469
	B	32	5,28	2,797	,494
zad.8	A	30	3,67	2,368	,432
	B	32	5,25	2,907	,514
zad.9	A	30	2,67	1,900	,347
	B	32	2,97	1,892	,334
zad.10	A	30	3,07	2,449	,447
	B	32	3,47	2,063	,365

Tabela 18 wyniki testu t-Studenta dla otrzymanych danych. Pogrubioną czcionką oznaczono wyniki istotne.

Tab.18. Test t równości średnich (założono równość wariancji)

Zad.	t	df	Istotność (dwustronna)	Różnica średnich	Błąd standardowy różnicy	95% przedział ufności dla różnicy średnich	
						Dolna granica	Górna granica
1.	-,262	60	,794	-,194	,740	-1,673	1,286
2.	-1,663	61	,101	-,950	,571	-2,092	,192
3.	-,409	60	,684	-,226	,552	-1,331	,879
4.	-1,379	61	,173	-,887	,643	-2,173	,399
5.	-1,187	60	,240	-,625	,526	-1,678	,428
6.	-,845	60	,402	-,435	,516	-1,467	,596
7.	-1,143	60	,258	-,781	,683	-2,148	,586
8.	-2,342	60	,023	-1,583	,676	-2,936	-,231
9.	-,627	60	,533	-,302	,482	-1,266	,662
10.	-,701	60	,486	-,402	,574	-1,550	,746

3.4 Użyteczność

Przeprowadzono również krótką ankietę, w której badani mogli zaznaczyć te spośród czterech zdań, z którymi się zgadzają. Wyniki zaprezentowane są na rysunkach 9-12.

Mógłbym (mogłabym) rozwiązywać takie zdanie:

- za każdym razem kiedy wysyłam e-mail — **15,6 %** (10 osób)



Rys.9.

- za każdym razem kiedy wysyłam e-mail, jeżeli dziesięciokrotnie zmniejszyłoby to liczbę spamu przychodzącego na moją skrzynkę — **43,8 %** (28 osób)



Rys.10.

- każdorazowo przy rejestrowaniu darmowego konta e-mail — **34,4 %** (22 osoby)



Rys.11.

- każdorazowo przy rejestrowaniu się na stronę związaną z poufnością danych (np. strony WWW banków), jeżeli znacznie zwiększyłoby to bezpieczeństwo transakcji — **65,6 %** (42 osoby)



Rys.12.

4 Dyskusja

Wyniki badań potwierdzają trzy postawione hipotezy badawcze.

Przede wszystkim, w przypadku skonstruowanych przez nas zadań testowych wystąpił efekt pozytywnego prymowania semantycznego (H1). Efekt ten zaobserwowano dla ośmiu, spośród dziesięciu, zadań testowych. Efektu semantycznego prymowania pozytywnego nie zaobserwowano w dwóch pierwszych zadaniach, co prawdopodobnie związane jest z koniecznością nabycia przez badanych pewnej wprawy w rozwiązywaniu zadań.

Dodatkowo różnica pomiędzy średnimi czasami rozwiązania zadania dla grupy eksperymentalnej i dla grupy kontrolnej jest istotna statystycznie dla przypadków kiedy zaobserwowano wystąpienie efektu prymowania. Tym samym możemy powiedzieć, że jesteśmy w stanie — na podstawie średniego czasu rozwiązania zadania — różnicować grupę eksperymentalną i grupę kontrolną (H2).

Co więcej, wyniki wskazują, że poprawa czasu rozwiązywania zadania (uzyskana dzięki pozytywnemu prymowaniu semantycznemu) nie wpływa w znaczący statystycznie sposób na obniżenie poprawności rozwiązań.

W przypadku jednego z zadań testowych potwierdziła się również trzecia hipoteza badawcza (dotycząca deklarowanej subiektywnej trudności zadań). Dla zadania ósmego deklarowana subiektywna trudność zadania była statystycznie niższa dla grupy eksperymentalnej.

Otrzymane wyniki wydają się być szczególnie obiecujące w przypadku zadania ósmego, jako dla tego zadania, które było różnicujące, jeżeli chodzi o deklarowaną subiektywną trudność rozwiązywanych zadań. W kolejnym badaniu należałoby wszystkie zadania skonstruować na wzór zadania ósmego (analogiczne ustawienia czasu wyświetlania prymy i maski oraz zniekształceń obrazu) i przetestować, czy otrzymamy również tak dobre rezultaty.

Ciekawe są również wstępne wyniki dotyczące postrzegania użytecznościowego systemu. Aż 65,6 % badanych zadeklarowało chęć rozwiązywania zaprojektowanych przez nas zadań *každorazowo* przy rejestrowaniu się na stronę związaną z poufnością danych. Duża liczba badanych byłaby też skłonna robić to *každorazowo* przy rejestrowaniu darmowego konta e-mail (43,4 %). Jeśli chodzi o chęć rozwiązywania zadania przy *každorazowym* wysyłaniu e-maila to zadeklarowało ją tylko 15,6 % badanych, jednakże liczba ta wzrasta do 43,8 % jeżeli zmniejszyłoby to dziesięciokrotnie liczbę spamu przychodzącego na skrzynkę e-mailową (a właśnie taki efekt przewidywany jest dla systemów CAPTCHA — por. [8]).

Literatura

- [1] Alario F.X., Segui J., Ferand L. Semantic and associative priming in picture naming. *The Quarterly Journal of Experimental Psychology*. 2000, 53A(3): 741–764.
- [2] Ahn L., Blum M., Hopper N.J., Langford J. CAPTCHA: Using Hard AI Problems For Security. Internet: <http://www.captcha.net>.
- [3] Ahn L., Blum M., Langford J. Telling Humans and Computers Apart Automatically. How Lazy Cryptographers do AI. Internet: <http://www.captcha.net>.
- [4] Baird H.S., Coates A.L., Fateman R.J. (2003). PessimPrint: a reverse Turing test. *International Journal on Document Analysis and Recognition* 5: 158–163.
- [5] Bergmair R., Katzenbeisser S. Towards Human Interactive Poofs in the Text-Domain. Using the Problem of Sense-Ambiguity for Security. Internet.
- [6] Bradford P. G., Wolowski M. (1994). A formalization of the Turing test. Tech. Rep. 339, 1994, Department of Computer Science, Indiana University. <http://www.cs.indiana.edu/pub/techreports/TR339.html>.
- [7] Chan N. Sound oriented CAPTCHA. Abstract. <http://www.capha.net>.
- [8] Chew M., Baird H. S. (2003). Baffle Text: a Human Interactive Proof. *Proceedings of the SPIE/IS&T Document Recognition & Retrieval Conf. X*. Santa Clara, January 22–23, 2003.
- [9] Goldreich O., Håstad J. (1997). On the Complexity of Interactive Proofs with Bounded Communication. Internet.
- [10] Goldwasser S., Micali S., Rackoff Ch. (1985). The Knowledge Complexity of Interactive Proof-Systems. Internet.
- [11] Goldwasser S., Spiser M. (1986). Private Coins versus Public Coins in Interactive Proof System. Internet.
- [12] Goldwasser S. Multi-Party Computations: Past and Present. Internet.

- [13] Mori G., Malik J. Recognizing Objects in Anversarial Clutter: Breaking a Visual CAPTHCHA.
www.cs.berkeley.edu/~mori/research/papers/mori_cvpr03.pdf
- [14] Naor M. (1996). Verification of a human in the loop or Identification via The Turing Test. Internet.
- [15] Nęcka E., Orzechowski J., Szymura B. **Psychologia poznawcza**. ACADEMICA & PWN, Warszawa 2006.
- [16] Newman M.H.A., Turing A.M., Jefferson B., Braithwaite R.B. (1952). Can automatic calculating machines be said to think?, broadcast discussion transmitted on BBC (14 and 23 Jan. 1952), the Turing Digital Archive (www.turingarchive.org), Contents of **AMT/B/6**.
- [17] Papadimitriou Ch. H. (2002). *Złożoność obliczeniowa*. Wydawnictwo Naukowo Techniczne: Warszawa.
- [18] Perea M, Rosa E., The effects of associative and semantic priming in the lexical decision task. *Psychological Research*. 2002, 66: 180–192.
- [19] Rui Y., Liu Z. (2004). ARTiFACIAL: Automated Reverse Turing test using FACIAL features. *Multimedia System* **9**: 493-502.
- [20] Shieber S.M. Turing test as Interactive Proof.
<http://www.eeccs.harward.edu/shieber/Biblio/Papers/turing-interactive-proof.pdf>.
- [21] Strelau J. (red.) (2000). *Psychologia*. Podręcznik akademicki. t. 2. Psychologia ogólna. Gdańskie Wydawnictwo Psychologiczne. Gdańsk.
- [22] Traiger S. (2000). Making the Right Identification in the Turing Test. *Mind and Machines* **10**: 561–572.
- [23] Turing A. M. (1950). Computing Machinery and Intelligence. *Mind*, vol. LIX, no 236/1950, s. 443-455.
- [24] Turing A. M. Can Digital Computers Think, the Turing Digital Archive (www.turingarchive.org), Contents of **AMT/B/5**.
- [25] Turing A. M. Intelligent Machinery, a Heretical Theory, the Turing Digital Archive (www.turingarchive.org), Contents of **AMT/B/4** and **B20**.

- [26] Turing A. M. Digital Computers Aplied to Games, **Faster than thought**, ed. B.V. BOWDEN, Bowden, London 1953; the Turing Digital Archive (www.turingarchive.org), Contents of **AMT/B/7**.
- [27] Watt S. Naive Psychology and the Inverted Turing Test. *Psycholoquy* 7(14)
<http://www.cogsci.esc.soton.ac.uk/cgi/psyc/newpsy?7.17>.

Załączniki

ZAŁĄCZNIK A.

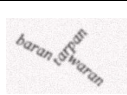
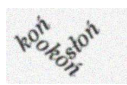


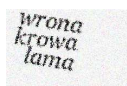


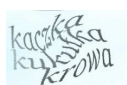
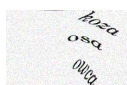

Zadania testowe

1. 
2. 
3. 
4. 
5. 
6. 
7. 
8. 
9. 
10. 

ZAŁĄCZNIK B.

Analiza zadań testowych programem AABYY Fine Reader PE 8.0

Poniższa tabela przedstawia analizę zadań testowych z wykorzystaniem programu AABYY Fine Reader PE 8.0. Kolumna „wejście” przedstawia obrazek, jaki program otrzymywał na wejściu, zaś kolumna „wyjście” przedstawia efekt prób odczytania słów znajdujących się na zadanym wejściu. Czas, jakiego potrzebował program zapisany jest w kolumnie „czas”.

	wejście	czas	wyjście
1.		00:67	A&.,**
2.		00:37	
3.		02:56	*4
4.		01:16	^marend®
5.		02:69	Prona krowa torna
6.		00:73	%s?
7.		01:64	█ m
8.		01:02	^{SX} AgbW
9.		00:96	
10.		00:51	k ^{ot} kruk

ZAŁĄCZNIK C.

Ankieta

Płeć: **K** / **M**

Wiek:

Wykształcenie:

Zawód:

Jak często korzystasz z Internetu?

- okazjonalnie
- często (przynajmniej raz w tygodniu)
- bardzo często (codziennie)

*Czy korzystasz z darmowych skrzynek e-mail? **tak** / **nie***

Mógłbym (mogłabym) rozwiązywać takie zdanie:

- za każdym razem kiedy wysyłam e-mail
 - za każdym razem kiedy wysyłam e-mail, jeżeli dziesięciokrotnie zmniejszyłoby to liczbę spamu przychodzącego na moją skrzynkę
 - każdorazowo przy rejestrowaniu darmowego konta e-mail
 - każdorazowo przy rejestrowaniu się na stronę związaną z poufnością danych (np. strony WWW banków), jeżeli znacznie zwiększyłoby to bezpieczeństwo transakcji
-